

## Feature Selection pada Azure Machine Learning untuk Prediksi Calon Mahasiswa Berprestasi

Hana Ariesta<sup>1</sup>, Maria Angela Kartawidjaja<sup>1</sup>

**ABSTRACT:** Universities can not only depend on digital marketing capabilities that are currently widely used but also utilize emerging technologies such as predictive analytics. This research attempts to make a predictive basis for analytic applications by utilizing Azure Machine Learning. In addition to utilizing machine learning capabilities, the author tries to use Feature Selection which is provided to obtain attributes that support research goals related to the world of education. This research aims to find relevant factors for outstanding student candidates for every student admission process. This research uses the Support Vector Machine Algorithm and comparison of the Feature Selection using Pearson Correlation Coefficient and Mutual Information. The results obtained using 10 attributes obtained the best results of 0.827 for accuracy and 0.831 for precision.

**KEYWORDS:** Azure, Machine Learning, Outstanding Students, Predictive, SVM

**ABSTRAK:** Institusi pendidikan tidak boleh hanya bergantung pada kemampuan digital marketing yang saat ini banyak digunakan namun juga memanfaatkan teknologi yang sedang berkembang seperti predictive analytic. Penelitian ini mencoba membuat dasar aplikasi predictive analytic dengan memanfaatkan Azure Machine Learning. Selain memanfaatkan kemampuan machine learning, penulis mencoba menggunakan Feature Selection yang disediakan untuk memperoleh atribut yang mendukung tujuan penelitian terkait dunia pendidikan. Penelitian ini bertujuan menemukan faktor yang relevan untuk calon mahasiswa berprestasi pada setiap penerimaan mahasiswa baru yang dilakukan. Penelitian ini menggunakan Algoritma Support Vector Machine serta perbandingan Feature Selection menggunakan Pearson Correlative Coefficient dan Mutual Information. Hasil yang diperoleh menggunakan 10 atribut memperoleh hasil terbaik 0.827 untuk accuracy dan 0.831 untuk precision.

**KATA KUNCI:** Azure, Machine Learning, Mahasiswa Berprestasi, Prediksi, SVM.

### PENDAHULUAN

Penerimaan mahasiswa baru pasti menjadi prioritas dalam setiap institusi pendidikan. Persaingan dalam bisnis pendidikan semakin ketat sehingga tidak mengherankan jika promosi yang semakin kreatif menjadi sesuatu yang penting. Tantangan dalam dunia marketing adalah bagaimana menekan atau mengurangi biaya untuk promosi dan mengoptimalkan sumber daya institusi yang tersedia dengan cara yang tepat dan bermanfaat. Untuk mengatasi hal tersebut, institusi dapat memanfaatkan teknologi yang sekarang ini sedang berkembang yaitu *learning analytic*. Teknologi ini sudah banyak digunakan dalam dunia bisnis seperti meningkatkan pertumbuhan investasi, mencari tren selera atau gaya hidup yang sedang



■ Gambar 1. Arsitektur Cloud Computing

<sup>1</sup> Magister Teknik Elektro, Universitas Katolik Indonesia Atma Jaya

terjadi dalam masyarakat, dan masih banyak lainnya. Namun dalam dunia edukasi teknologi ini lebih digunakan untuk memonitor atau mengukur kemajuan siswa. Jika dilihat lebih jauh, peluang memanfaatkan teknologi *learning analytic* untuk memprediksi keberhasilan atau menentukan strategi marketing yang memungkinkan.

Untuk mengembangkan dan mempelajari mengenai *learning analytic* membutuhkan penggunaan *machine learning* untuk membentuk model prediksi dari data yang sudah diolah. Berdasarkan kemampuan dari *machine learning*, penelitian yang akan dibuat ini berpotensi meningkatkan kualitas pengolahan data yang saat ini masih kurang optimal. Contoh sumber data yang dapat diolah antara lain dari hasil *history* penelusuran situs web resmi universitas, data pertanyaan seputar admisi, dan juga data akademis seperti data nilai rapor dari calon mahasiswa yang mendaftar, data indeks prestasi mahasiswa dan lama studi mahasiswa. Berdasarkan data tersebut, algoritma *machine learning* memungkinkan untuk menghasilkan model prediksi calon mahasiswa berprestasi yang dapat digunakan untuk menambah jumlah mahasiswa baru pada universitas.

Tujuan penelitian ini untuk membangun sebuah model prediksi data serta memperoleh faktor relevan yang berpengaruh terhadap data calon mahasiswa yang mendaftar penerimaan mahasiswa baru melalui jalur bebas tes dan berpotensi memiliki performa yang baik di tahun pertama kuliah. Dari hasil tersebut dapat menambah suatu bentuk digital marketing baru untuk universitas. Tim marketing nantinya dapat langsung menggunakan data untuk menawarkan potongan SPP atau beasiswa agar tertarik untuk mendaftar ke universitas

Pengolahan data yang besar membutuhkan alat bantu untuk menganalisis data menjadi pengetahuan yang mendukung kemajuan perguruan tinggi di masa yang akan datang. Teknologi yang banyak digunakan saat ini mengacu pada *cloud computing*. Pengguna dapat memanfaatkan hasil pengolahan, analisis dan menyimpan data pada satu sumber yang sama. Teknologi *cloud computing* dibagi menjadi tiga tingkatan yaitu: *infrastructure as a service* (IaaS), *platform as a service* (PaaS) dan *software as a service* (SaaS) seperti ditunjukkan pada Gambar 1. Sudah banyak penyedia layanan teknologi *cloud computing* seperti Amazon Web Services (AWS), *Google Cloud Services*, Microsoft Azure dan lainnya [8].

## KAJIAN PUSTAKA

### **MACHINE LEARNING**

*Machine learning* merupakan salah satu sub bagian pembelajaran dari kecerdasan buatan yang memungkinkan sistem memiliki kemampuan untuk “belajar secara otomatis”. *Machine learning* dapat meningkatkan kemampuan berdasarkan pengalaman tanpa perlu diprogram secara eksplisit atau tanpa perlu intervensi manusia. *Machine learning* harus terlebih dahulu diberikan data untuk dapat dipelajarinya.

*Machine learning* memiliki sejumlah metode pembelajaran. Secara umum, *machine learning* diklasifikasikan ke dalam empat kategori, yaitu *Supervised Learning*, *Unsupervised Learning*, *Semi Supervised Learning* dan *Reinforcement Learning*.

Penelitian ini akan menggunakan *Supervised Learning*. Dalam penelitian ini proses yang dilakukan dibagi menjadi beberapa tahap, yaitu tahap *data preprocessing*, tahap pembersihan data, tahap penentuan atribut (*feature*), tahap klasifikasi, dan tahap evaluasi.

■ **Tabel 1** *Confusion Matrix* [7]

	Prediksi	
	Positive	Negative
Positive	True Positive (TP)	False Negative (FN)
Negative	False Positive (FP)	False Negative (FN)

## 1. TAHAP DATA PREPROCESSING

Pada tahap ini, data yang akan digunakan untuk penelitian dikumpulkan dalam satu sumber sehingga memudahkan dalam menghubungkan data antar tabel yang berbeda. Sumber data yang digunakan dalam penelitian ini didapat melalui data mahasiswa yang ada di Biro Administrasi Akademik Universitas Katolik Atma Jaya Jakarta pada angkatan tahun 2011 hingga 2013 untuk Fakultas Teknik, Fakultas Kedokteran dan Fakultas Teknobiologi.

## 2. PEMBERSIHAN DATA

Dalam proses ini akan diperiksa kembali jika terdapat data-data yang belum lengkap atau tidak akurat lebih baik tidak digunakan. Pembersihan data ini akan mempengaruhi performa perhitungan sesuai algoritma yang digunakan karena mengurangi kompleksitas. Data yang mempunyai nilai sama atau redundan akan dihapus. Begitu pula data yang tidak konsisten dan tidak lengkap (*missing value*). Pada penelitian ini, data yang tidak lengkap akan diisi dengan nilai rata-rata dari kolom data yang kosong tersebut.

## 3. TAHAP PENENTUAN ATRIBUT (*FEATURE*)

Penelitian ini menggunakan salah satu metode *feature selection* dalam *Azure Machine Learning* yaitu *Pearson Correlation Coefficient* dan *Mutual Information* dalam membantu memilih atribut yang mempunyai korelatif terkait tujuan penelitian.

## 4. TAHAP KLASIFIKASI

Penelitian ini menggunakan model klasifikasi *supervised learning* untuk mengevaluasi performa dari model machine learning ini karena menggunakan data yang berlabel. Untuk algoritma yang digunakan adalah *Support Vector Machine* (SVM).

## 5. TAHAP EVALUASI

Dalam tahap ini dilakukan pengujian untuk mencari dan membandingkan jumlah atribut sesuai skenario yang akan disusun menggunakan algoritma yang sudah dipilih yaitu SVM.

*Confusion matrix* adalah suatu metode yang biasanya digunakan untuk melakukan perhitungan akurasi pada konsep *data mining*. Hasil evaluasi calon mahasiswa berprestasi ini dapat dilihat menggunakan *confusion matrix* seperti pada *paper* Sembiring dan timnya (Sembiring, dkk., 2011). Tingkat akurasi dan tingkat kesalahan prediksi menunjukkan performa algoritma klasifikasi. *Confusion matrix* mempunyai empat bagian, yaitu *true positive*, *true negative*, *false positive* dan *false negative* seperti pada Tabel 1.

Akurasi merupakan jumlah prediksi benar dari semua data yang diprediksi yang dapat dihitung menggunakan rumus:

$$\text{Akurasi (\%)} = \frac{TP+TN}{TP+FP+FN+TN} \dots\dots\dots(1)$$

Performa keakurasian diklasifikasikan menjadi beberapa kelompok yaitu:

- a) nilai AUC 0.90 – 1.00 = Unggul
- b) nilai AUC 0.80 – 0.90 = Baik
- c) nilai AUC 0.70 – 0.80 = Cukup
- d) nilai AUC 0.60 – 0.70 = Kurang
- e) nilai AUC 0.50 – 0.60 = Gagal

## **MICROSOFT AZURE**

Microsoft Azure adalah suatu platform yang ditawarkan oleh perusahaan Microsoft untuk mengolah dan menganalisis data terutama data yang banyak melalui jaringan global. Untuk mengantisipasi meningkatnya data maka Microsoft menawarkan layanan yang disebut dengan Microsoft Azure. Untuk menangani analisis prediktif (*predictive analysis*), layanan yang ditawarkan adalah *Azure Machine learning*. Layanan ini tidak hanya menyediakan alat untuk memodelkan analisis prediktif, namun juga menyediakan layanan yang memungkinkan pengguna untuk mengelola model prediksi yang sudah dibangun.

## **FEATURE SELECTION**

*Feature selection* digunakan dalam tahap *data preprocessing* untuk mengetahui fitur yang paling signifikan dan intrinsik. Dalam Microsoft Azure, terdapat beberapa metode untuk melakukan *feature selection* yaitu: *Chi-square*, *Pearson Correlation Coefficient*, dan *Mutual Information* seperti dipaparkan dalam jurnal IEEE oleh Punlumjeak [10]. Pada penelitian ini akan membandingkan antara *Pearson Correlation Coefficient* dan *Mutual Information*.

## **PEARSON CORRELATION COEFFICIENT**

*Pearson Correlation Coefficient* adalah pengukuran statistik, yang mengukur kekuatan hubungan antara distribusi nilai fitur  $x$  dan label kelas  $c$ . *Pearson Correlation Coefficient* didefinisikan sebagai Persamaan sebagai berikut:

$$\rho(X,C) = \frac{E(XC) - E(X)E(C)}{\sqrt{\sigma^2(X)\sigma^2(C)}} \dots\dots\dots (2)$$

$$\rho(X,C) = \frac{\sum_i (x_i - \bar{x}_i)(c_i - \bar{c}_i)}{\sqrt{\sum_i (x_i - \bar{x}_i)^2 \sum_j (c_j - \bar{c}_j)^2}} \dots\dots\dots (3)$$

$\rho(X, C) = \pm 1$ , jika nilai fitur  $x$  bergantung linier. Di sisi lain, jika tidak berkorelasi dianggap nol.

## **MUTUAL INFORMATION**

Dua variabel acak pada *Mutual Information* adalah variabel yang saling bergantung jika dalam teori informasi. Dua variabel acak diskrit  $X$  dan  $Y$  dalam *Mutual Information* dapat didefinisikan pada Persamaan (4):

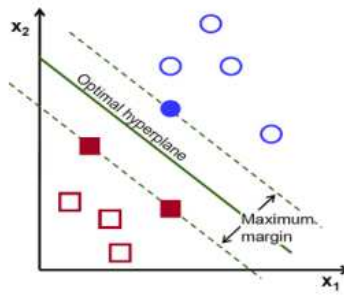
$$I(X;Y) = \sum_{x,y} P_{xy}(x,y) \log \frac{P_{xy}(x,y)}{P_x(x)P_y(y)} \dots\dots\dots (4)$$

## **SUPPORT VECTOR MACHINE (SVM)**

*Support Vector Machine* adalah suatu teknik untuk melakukan prediksi, baik dalam kasus klasifikasi maupun regresi. SVM memiliki prinsip dasar *linear classifier* yaitu kasus klasifikasi yang secara linier dapat dipisahkan, namun SVM telah dikembangkan agar dapat bekerja pada problem non-linier dengan memasukkan konsep kernel pada ruang kerja berdimensi tinggi. Pada ruang berdimensi tinggi, akan dicari *hyperplane* yang dapat memaksimalkan jarak (margin) antara kelas data. *Hyperplane* klasifikasi linier SVM dinotasikan pada persamaan 5:

$$F(x) = W^T x + b \dots\dots\dots (5)$$

*Support Vector Machine* (SVM) didefinisikan sebagai seperangkat metode pembelajaran terkait yang menganalisis data dan mengenali pola, yang kemudian digunakan untuk klasifikasi dan analisis regresi. SVM mengambil satu set data input dan memprediksi untuk setiap masukan



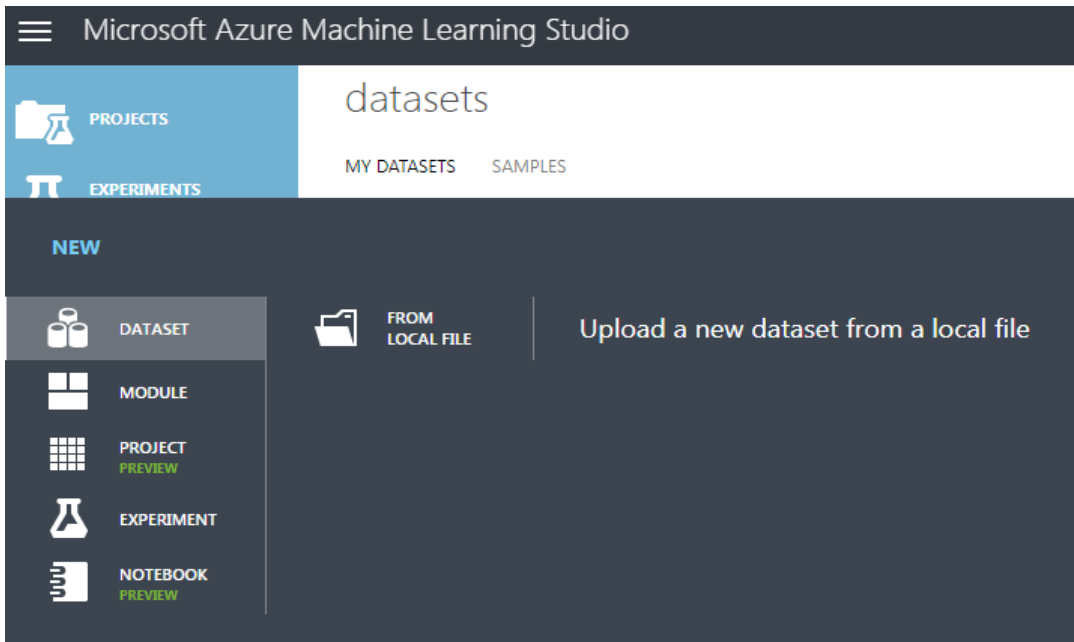
■ **Gambar 2.** *SVM Classification with optimal*

yang diberikan, yang berasal dari dua kelas yang kemudian di klasifikasikan dengan mencari nilai *hyperplane* terbaik [6].

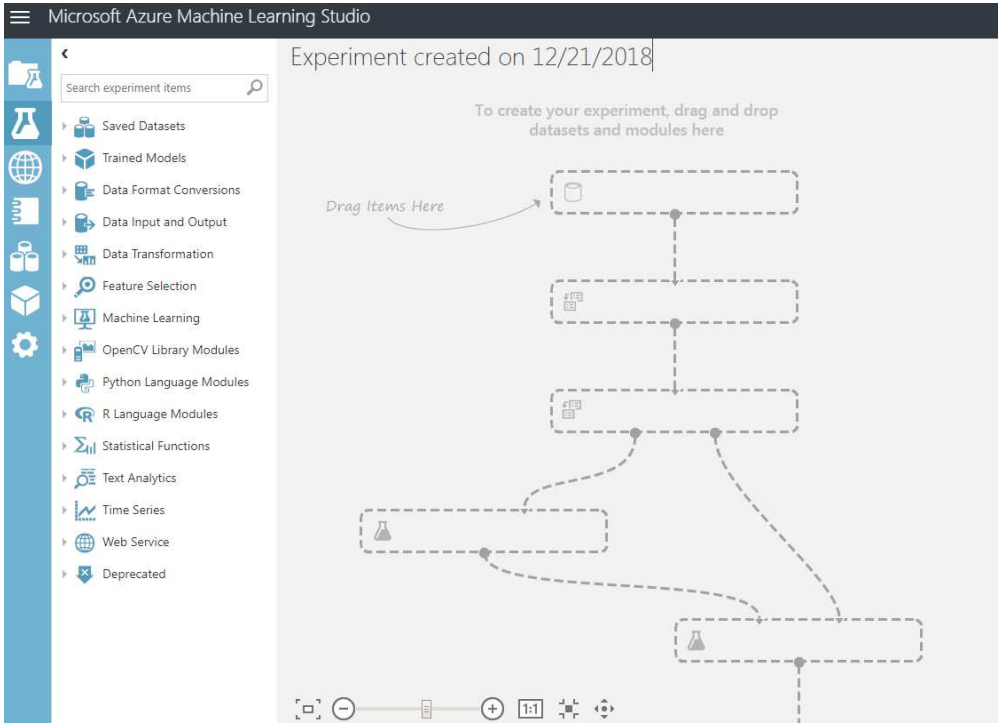
## METODOLOGI

Tahap awal penelitian pada platform Azure, dibuat sebuah percobaan baru pada menu *Experiment* dalam *Machine Learning Studio* seperti ditunjukkan pada Gambar 3. Sebelum membuat sebuah percobaan, disiapkan suatu *dataset* atau data yang akan digunakan pada percobaan terlebih dahulu. *Dataset* dapat diunggah pada menu *Dataset* dengan menyesuaikan format yang didukung oleh *Microsoft Azure Machine Learning*. Pada penelitian ini, penulis menggunakan format *.csv*. Setelah *dataset* siap, langkah selanjutnya adalah memilih dataset mana yang akan digunakan. Tampilan halaman utama pada Menu *Experiment* Azure Machine Learning dapat dilihat pada Gambar 4.

Tahap berikutnya membagi data menjadi *data train* dan *data model*. Variabel perbandingan *data train* dan *data model* dapat diatur sesuai kebutuhan. Pada penelitian ini, penulis membuat perbandingan data latih dan data uji sebesar 75% : 25%. Dari total data awal



■ **Gambar 3.** Menu *Dataset* pada *Azure Machine Learning*



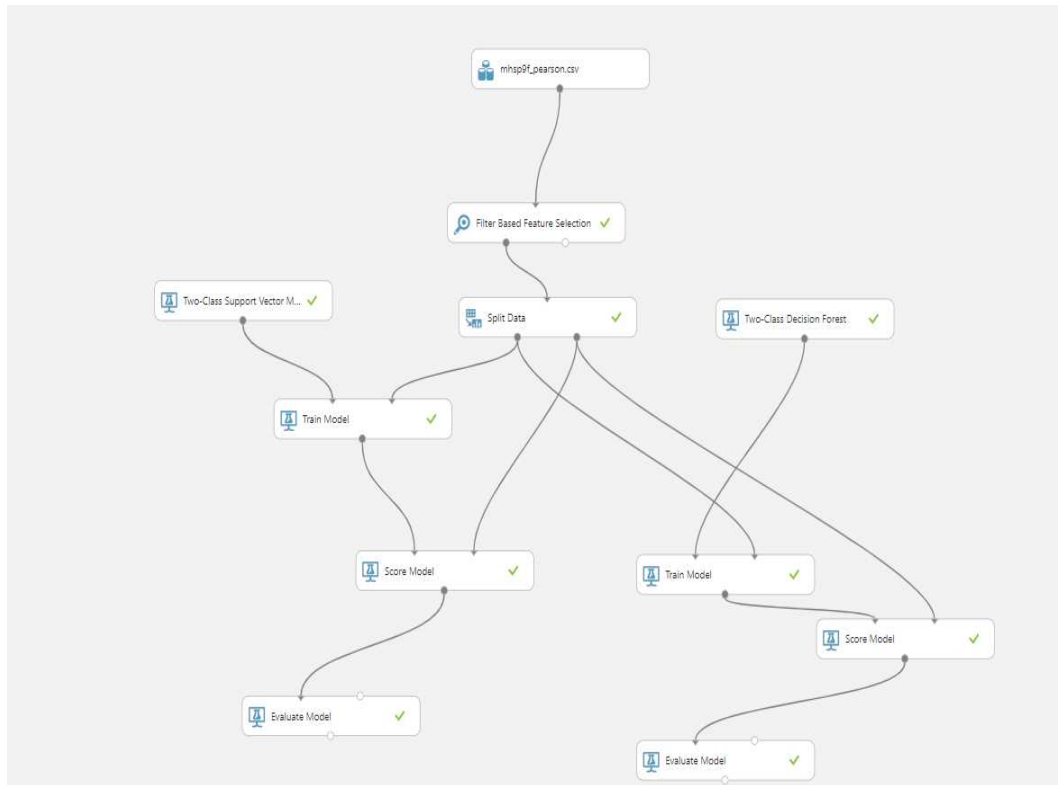
Gambar 4. Tampilan Halaman Utama membuat Percobaan

■ Tabel 2 Daftar Atribut

No.	Nama Atribut	Deskripsi
1	GPA 1	Rata-rata GPA pada semester 1 & 2
2	GPA 2	Rata-rata GPA semester 3 & 4
3	CGPA	GPA saat lulus.
4	MAT	Rata-rata nilai Matematika Kelas 10 & 11
5	KIM	Rata-rata nilai Kimia Kelas 10 & 11
6	FIS	Rata-rata nilai Fisika Kelas 10 & 11
7	BIO	Rata-rata nilai Biologi Kelas 10 & 11
8	Ekskul	Aktif di Ekstrakurikuler
9	OSIS	Aktif di OSIS
10	Scholarship	Penerima beasiswa

■ Tabel 3 Label Data

Label
Lulus Tepat Waktu
Tidak Lulus Tepat Waktu



■ **Gambar 5.** Diagram Alur Percobaan pada Azure Machine Learning

sebanyak kurang lebih 800 data, maka dibagi menjadi 600 data sebagai data latih (*train data*) dan 200 data sebagai data uji (*testing data*). Untuk atribut, penelitian ini menggunakan 10 atribut dan 2 label dengan deskripsi masing-masing seperti ditunjukkan dalam Tabel 2 dan Tabel 3.

Pada penelitian ini, label yang digunakan dibagi menjadi dua kategori yaitu lulus tepat waktu dan tidak lulus tepat waktu. Untuk lulus waktu perhitungannya diambil dari tahun masuk sampai tahun lulus tidak lebih dari 4 tahun atau 8 semester dan sebaliknya untuk kategori tidak lulus tepat waktu diambil dari hasil yang lulus lebih dari 4 tahun atau 8 semester.

Pada penelitian ini, algoritma yang digunakan adalah *Support Vector Machine* (SVM) dengan menggunakan metode *Pearson Correlation Coefficient* dan *Mutual Information* pada *feature selection* yang dilakukan pada proses *data preprocessing*. Diagram Alur percobaan yang dibuat pada Azure Machine Learning Studio dapat dilihat pada Gambar 5. Percobaan ini akan membandingkan kinerja algoritma Support Vector Machine dengan jumlah atribut yang berbeda yaitu 6 atribut, 8 atribut, dan 10 atribut. Penelitian ini menggunakan enam skenario tampak pada tabel 4.

### HASIL PENELITIAN

Dari penelitian yang sudah dilakukan sesuai skenario dihasilkan *confusion matrix* yang tampak pada tabel 5 hingga tabel 10. Data *confusion matrix* di atas didapat melalui *data testing*. Dari *confusion matrix* seluruh skenario, maka didapatkan daftar performa akurasi data dari model pengolahan data menggunakan algoritma terpilih sebagai berikut:

Penghitungan *Precision* diambil dari: *True Positive* (TP) dibagi dengan hasil penjumlahan antara *True Positive* (TP) dan *False Positif* (FP). Melihat dari hasil pengukuran *accuracy*, *precision*, dan AUC terhadap 200 data *testing* dengan total 10 atribut menggunakan

■ **Tabel 4 Daftar Skenario**

No.	Skenario	<i>Feature Selection &amp; Jumlah Atribut</i>
1	Skenario 1	<i>Pearson – 6 Atribut</i>
2	Skenario 2	<i>MI – 6 Atribut</i>
3	Skenario 3	<i>Pearson – 8 Atribut</i>
4	Skenario 4	<i>MI – 8 Atribut</i>
5	Skenario 5	<i>Pearson – 10 Atribut (All)</i>
6	Skenario 6	<i>MI – 10 Atribut (All)</i>

■ **Tabel 5. Confusion Matrix Skenario 1**

		Prediksi	
		<i>Positive</i>	<i>Negative</i>
Aktual	<i>True</i>	103	21
	<i>False</i>	21	17

■ **Tabel 6 Confusion Matrix Skenario 2**

		Prediksi	
		<i>Positive</i>	<i>Negative</i>
Aktual	<i>True</i>	96	16
	<i>False</i>	23	90

■ **Tabel 7 Confusion Matrix Skenario 3**

		Prediksi	
		<i>Positive</i>	<i>Negative</i>
Aktual	<i>True</i>	102	22
	<i>False</i>	22	76

■ **Tabel 8 Confusion Matrix Skenario 4**

		Prediksi	
		<i>Positive</i>	<i>Negative</i>
Aktual	<i>True</i>	97	15
	<i>False</i>	24	89

■ **Tabel 9 Confusion Matrix Skenario 5**

		Prediksi	
		<i>Positive</i>	<i>Negative</i>
Aktual	<i>True</i>	96	16
	<i>False</i>	23	90

■ **Tabel 10 Confusion Matrix Skenario 6**

		Prediksi	
		<i>Positive</i>	<i>Negative</i>
Aktual	<i>True</i>	102	22
	<i>False</i>	22	76



■ Tabel 11. *Overall Accuracy*

Skenario	<i>Accuracy</i>	<i>Precision</i>	AUC
1	0.811	0.831	0.868
2	0.827	0.807	0.898
3	0.802	0.823	0.866
4	0.827	0.802	0.896
5	0.827	0.807	0.897
6	0.802	0.823	0.866

algoritma *Support Vector Machine* terlihat bahwa sistem mencapai performa optimal seperti yang diharapkan. Pada skenario pertama menggunakan *Pearson Correlation Coefficient* dengan 6 atribut mempunyai hasil *accuracy* sebesar 0.811 dengan *precision* 0.831. Skenario kedua menggunakan *Mutual Information* dan 6 atribut mendapatkan hasil *accuracy* sebesar 0.827, *precision* 0.807 dengan AUC sebesar 0.898. Untuk skenario ketiga menggunakan *Pearson Correlation Coefficient* dan 8 atribut, mempunyai nilai *accuracy* lebih rendah dari skenario pertama dan kedua yaitu 0.802 dan *precision* 0.823 serta AUC 0.866. Untuk skenario keempat menggunakan *Mutual Information* dan 8 atribut mempunyai hasil *accuracy* sebesar 0.827 *precision* lebih rendah dari skenario sebelumnya yaitu 0.802 dan AUC sebesar 0.896. Pada skenario

kelima menggunakan *Pearson Correlation Coefficient* dan semua atribut menghasilkan *accuracy* sebesar 0.827, *precision* 0.807 dengan AUC sebesar 0.897 sedangkan untuk skenario keenam hasil *accuracy* sebesar 0.802, *precision* 0.823 dengan AUC sebesar 0.866. Dari keseluruhan skenario tersebut dapat dilihat hasil *accuracy* terbaik didapat pada skenario kedua, ketiga dan keempat. Namun hasil *precision* yang terbaik justru pada skenario pertama. Lain halnya untuk tiga hasil AUC terbaik diperoleh melalui skenario kedua lalu kelima dan keempat.

## KESIMPULAN

Melihat hasil penelitian yang sudah dilakukan dapat diambil kesimpulan sebagai berikut :

1. Penggunaan algoritma SVM dengan memilih metode *Pearson Correlation Coefficient* maupun *Mutual Information* pada *feature selection* sudah sesuai dengan hasil *accuracy* optimal yang diharapkan yaitu lebih dari 0.8. Begitu juga dengan hasil *precision* serta AUC. Hal ini membuktikan penggunaan *Feature Selection* pada Azure Machine Learning membawa dampak positif dalam penelitian klasifikasi seperti percobaan ini.
2. Faktor-faktor yang relevan dengan prestasi mahasiswa pada tahun pertama perkuliahan berdasarkan hasil *Precision* terbaik adalah beasiswa, nilai kimia, nilai matematika, nilai biologi, ekstrakurikuler dan nilai fisika. Hasil tersebut sesuai urutan nilai bobot masing-masing atribut yang didapat pada percobaan.
3. Hasil dari keseluruhan percobaan yang telah dilakukan belum mencapai nilai maksimum yaitu diatas 0.9 seperti yang diharapkan karena keterbatasan jumlah dan kualitas data yang digunakan.

Penelitian menggunakan Microsoft Azure ini masih mungkin dikembangkan terutama dengan memilih menggunakan algoritma Machine Learning yang lain seperti *Logistic Regression*, *Decision Forest* dan lainnya dengan jumlah data uji yang lebih banyak dan berkualitas. Selain itu disarankan untuk menambah jumlah atribut sehingga hasil akurasi yang diperoleh dapat lebih baik lagi.

## **DAFTAR REFERENSI**

- [1] Lusi Ariyani, 2016, “Kajian Penerapan Model C45, Support Vector Machine (SVM), dan Neural Network dalam Prediksi Kenaikan Kelas”, *Faktor Exacta* 9(1): 72-86, 2016. ISSN: 1979-276x.
- [2] Fahriah, Sirli dan Heru Agus Santoso. Penerapan Algoritma ID3 untuk perediksi Minat Studi Mahasiswa Teknik Informatika Pada Universitas Dian Nuswantoro.
- [3] Ahmad, Iftikhar. Basher, Mohammad and Iqbal, Muhammad Javed. 2018, “Performance Comparison on Support Vector Machine, Random Forest, and Extreme Learning Machine for Intrusion Detection”, *IEEE Access* Volume 6, 2018.
- [4] Li, Jundong and Liu, Huan, 2017, “Challenges of Feature Selection for Big Data Analytics”. *IEEE Computer Society*, March/April 2017.
- [5] Zollanvari, Amin. Kizilirmak, Refik Caglar, Kho, Yau Hee and Torrano, Daniel Hernandez, 2018, “Predicting Students’ GPA and Developing Intervention Strategies Based on Self-Regulatory Learning Behaviors”. *IEEE Access*, Volume 5, 2017.
- [6] Yin, Y., Han, D., & Cai, Z. (2011). Explore Data Classification Algorithm Based on SVM and PSO for Education Decision. *Journal of Convergence Information Technology*, 6(10), 122–128.
- [7] Betha Nurina Sari, 2017, “Prediksi Performa Akademik Siswa Pada Pelajaran Matematika Menggunakan Bayesian Networks dan Algoritma Klasifikasi Machine Learning”, *KNPMP II*, Universitas Muhammadiyah Surakarta. ISSN: 2502-6526.
- [8] <http://aryatryasthana.blogspot.com/2015/10/tugas-v-cloud-computing-e-commerce-dan.html> diakses pada 28 September 2018.
- [9] Sembiring, S., Zarlis, M., Hartama, D., Ramliana, S., & Wani, E. (2011). Prediction of Student Academic Performance By an Application of Data Mining Techniques. *International Conference on Management and Artificial Intelligence (IPEDR)*, 6, 110–114.
- [10] Punlumjeak, Wattana. Rachburee, Nachirat and Arunrerk, Jedsada. “Big Data Analytics: Student Performance Prediction Using Feature Selection and Machine Learning on Microsoft Azure Platform”, *Journal of Telecommunication, Electronic and Computer Engineering*. Vol.9 No.1-4.